was crucial to the discrimination between transmitting and normal males. Yunis and Soreng[16] first showed that caffeine, which has synergistic effects with a variety of DNA damaging agents, enhances expression of common fragile sites and fragile X when used in conjunction with fluorodeoxyuridine (FUdR). Although we and others[18-19] have failed to replicate caffeine's enhancing effect on fragile X expression in lymphocytes or lymphoblasts, it has consistently increased expression in hybrid cells[6,7]. This difference may be due to the fact that caffeine's effect on fragile X expression appears to be indirect: it inhibits the mitotic delay usually associated with DNA damage, thereby reducing the time available for DNA replication, repair and chromosome condensation[7]. The effect may vary among cell types with different cell cycle characteristics, particularly different lengths of $G_2$.

The high prevalence and unusual segregation patterns observed in fragile X syndrome suggest that it is not a classical X-linked mutation[3,12,20]. The single most unusual characteristic of its inheritance is its variable penetrance within pedigrees. For example, the penetrance of mental retardation is significantly lower in offspring of mothers of transmitting males than in offspring of daughters of transmitting males[12]. This observation has been referred to as the 'Sherman paradox'[21] and must be explained by any model addressing inheritance of fragile X. The unique ability of the hybrid system to induce fragile X expression in transmitting and normal males and to demonstrate a continuum of cytogenetic expression (low in normal males, higher in transmitting males, highest in affected males) argues strongly against the involvement of autosomal modifying loci in this disorder. Instead, the presence of a common fragile site at Xq27 is consistent with the multi-step mutation models of Pembrey *et al.*[20] and Nussbaum *et al.*[22].

We speculate that recombination between two normal X chromosomes at the Xq27 common fragile site region may be the mechanism for an initiation event, producing a clinically normal transmitting male or carrier female. Segregation data demonstrates that progression to the full XLMR mutation occurs only in females[20] (who have two X chromosomes), also suggesting that meiotic recombination is involved. These recombination events may produce variation in the length or copy number of a DNA sequence at Xq27, which at some threshold level disrupts the function of a neighbouring gene and produces XLMR. Quantitative variation in the size of the fragile site DNA region may also explain the variable recombination distances between polymorphic DNA markers observed among fragile X families[23]. Thus, our data suggest that a normal DNA sequence at Xq27 may be altered to produce a continuous quantitative variation in fragile site DNA resulting in varying degrees of cytogenetic expression and a threshold for clinical manifestation of the mutation.

We thank Drs P. A. Jacobs and W. T. Brown for providing lymphoblastoid cell lines on transmitting males. This work was supported by NIH RO1 HD20227, the Robert J. Kleberg, Jr Center for Human Genetics, and the Howard Hughes Medical Institute. R.L.N. is an Associate Investigator in the Howard Hughes Medical Institute.

Received 30 May; accepted 26 August 1986.

1. Sutherland, G. R., Hecht, F., Mulley, J. C., Glover, T. W. & Hecht, B. K. (eds) *Fragile Sites on Human Chromosomes*, 304 (Oxford University Press, New York, 1985).
2. Froster-Iskenius, U., McGillivray, B. C., Dill, F. J., Hall, J. G. & Herbst, D. S. *Am. J. med. Genet.* **232**, 619-632 (1986).
3. Howard-Peebles, P. N. & Friedman, J. M. *Am. J. hum. Genet.* **37**, 956-964 (1985).
4. Nussbaum, R. L., Airhart, S. D. & Ledbetter, D. H. *Hum. Genet.* **64**, 148-150 (1983).
5. Nussbaum, R. L., Walmsley, R. M., Lesko, J. G., Airhart, S. D. & Ledbetter, D. H. *Am. J. hum. Genet.* **37**, 1192-1205 (1985).
6. Ledbetter, D. H., Airhart, S. D. & Nussbaum, R. L. *Am. J. med. Genet.* **23**, 429-444 (1986).
7. Ledbetter, D. H., Airhart, S. D. & Nussbaum, R. L. *Am. J. med. Genet.* **23**, 445-456 (1986).
8. Warren, S. T. & Davidson, R. L. *Somat. Cell molec. Genet.* **10**, 409-413 (1984).
9. Rhoads, F. A., Oglesby, A. C., Mayer, M. & Jacobs, P. A. *Am. J. med. Genet.* **12**, 205-217 (1982).
10. Brown, W. T., Gross, A. C., Chan, C. B. & Jenkins, E. C. *Hum. Genet.* **71**, 11-18 (1985).
11. Steinbach, P. *Hum. Genet.* **72**, 248-252 (1986).
12. Sherman, S. L., Jacobs, P. A., Morton, N. E., Froster-Iskenius, U. & Howard-Peebles, P. N. *Hum. Genet.* **69**, 289-299 (1985).
13. Jenkins, E. C., Brown, W. T., Brooks, J., Duncan, C. J., Sanz, M. M. *et al. Expl Cell Biol.* **54**, 40-48 (1986).
14. Berger, R., Bloomfield, C. D. & Sutherland, G. R. *Cytogenet. Cell Genet.* **40**, 490-535 (1985).
15. Glover, T. W. in *Fragile Sites on Human Chromosomes* (eds Sutherland, G. R., Hecht, F., Mulley, J. C., Glover, T. W., Hecht, B. K.) 80-94 (Oxford University Press, New York, 1985).
16. Yunis, J. J. & Soreng, A. L. *Science* **226**, 1199-1204 (1984).
17. Daniel, A. *Am. J. med. Genet.* **23**, 419-427 (1986).
18. Glover, T. W., Coyle-Morris, J. & Morgan, R. *Cancer Genet. Cytogenet.* **19**, 141-150 (1986).
19. Abruzzo, M. A., Pettay, D., Mayer, M. & Jacobs, P. A. *Hum. Genet.* **73**, 20-22 (1986).
20. Pembrey, M. E., Winter, R. M. & Davies, K. E. *Am. J. med. Genet.* **21**, 709-718 (1985).
21. Opitz, J. M. *Am. J. med. Genet.* **23**, 1-10 (1986).
22. Nussbaum, R. L., Airhart, S. D. & Ledbetter, D. H. *Am. J. med. Genet.* **23**, 715-722 (1986).
23. Brown, W. T., Gross, A. C., Chan, C. B. & Jenkins, E. C. *Am. J. med. Genet.* **23**, 643-664 (1986).

# Analysis of enzymatically amplified $\beta$-globin and HLA-DQ$\alpha$ DNA with allele-specific oligonucleotide probes

**Randall K. Saiki, Teodorica L. Bugawan, Glenn T. Horn, Kary B. Mullis & Henry A. Erlich**

Department of Human Genetics, Cetus Corporation, 1400 Fifty-Third Street, Emeryville, California 94608, USA

**Allelic sequence variation has been analysed by synthetic oligonucleotide hybridization probes which can detect single base substitutions in human genomic DNA[1-3]. An allele-specific oligonucleotide (ASO) will only anneal to sequences that match it perfectly, a single mismatch being sufficient to prevent hybridization under appropriate conditions[4]. To improve the sensitivity, specificity and simplicity of this approach, we used the polymerase chain reaction (PCR) procedure[5] to enzymatically amplify a specific segment of the $\beta$-globin or HLA-DQ$\alpha$ gene in human genomic DNA before hybridization with ASOs. This *in vitro* amplification method, which produces a $>10^5$-fold increase in the amount of target sequence, permits the analysis of allelic variation with as little as 1 ng of genomic DNA and the use of a simple 'dot blot' for probe hybridization. As a further simplification, PCR amplification has been performed directly on crude cell lysates, eliminating the need for DNA purification.**

$^{32}$P-labelled ASO probes have been used in the analysis of a number of genetic diseases caused by single base mutations, including sickle cell anaemia[1], $\alpha$-1 anti-trypsin deficiency[2] and $\beta$-thalassemia[3]. Although the ASO method is a powerful diagnostic tool, it requires 5 to 10 $\mu$g of genomic DNA, gel-purified high specific activity probes and several days of autoradiographic exposure. The short oligomer probes (usually 19 bases long) also hybridize to other genomic sequences so that restriction endonuclease digestion and gel electrophoresis are required to separate the target sequence from the bulk of the genomic DNA. In order to develop a simple, sensitive method, we chose the sickle cell anaemia and haemoglobin C mutations in the sixth codon of the $\beta$-globin gene[6] as a model system for genetic diagnosis. We used ASOs specific for the normal ($\beta^A$), sickle cell ($\beta^S$), and haemoglobin C ($\beta^C$) sequences as probes to detect these alleles in PCR-amplified genomic samples. Figure 1 shows the sequence of the PCR primers (PC03 and PC04), the ASO probes (19A, 19S and 19C), and their relationship to the $\beta$-globin gene. The PCR primers, both 20 bases in length, anneal to opposite strands of the genomic DNA flanking the target sequence[7]. These primers are positioned so that the polymerase-catalysed, template-directed extension product of one primer can serve as a template strand for the other. In this configuration, repeated cycles of denaturation, primer annealing, and primer extension, result in the exponential accumulation of a 110-base pair (bp) $\beta$-globin fragment, with a length defined by the distance between the 5′ ends of the PCR primers. For example, 20 cycles of PCR amplification with PC03 and PC04 as primers produce a 220,000-fold increase in the amount of $\beta$-globin target

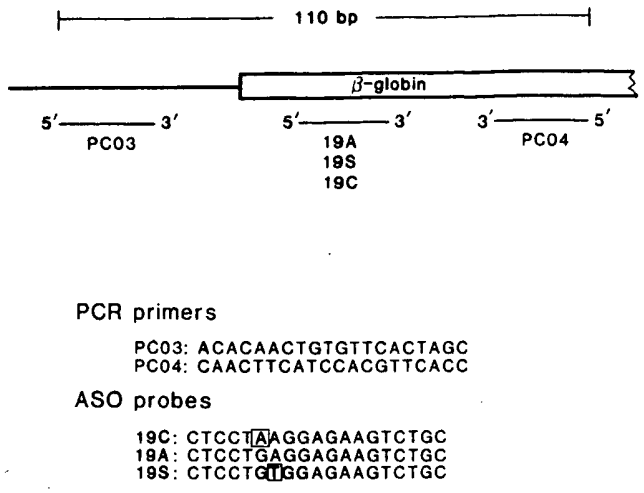


**Fig. 1** Sequence of the PCR primers and ASO probes and their relationship to the β-globin gene. The two PCR primers, PC03 and PC04, are 20 bases long and are complementary to the (−) and (+) strands, respectively. The three ASO probes, 19A, 19S, and 19C, are 19 bases long and span codons 4–17 of the gene[8]. 19A is complementary to the normal β-globin allele ($\beta^A$), 19S to the sickle cell allele ($\beta^S$), and 19C to the haemoglobin C allele ($\beta^C$). Boxes, single base mutations of $\beta^S$ and $\beta^C$.

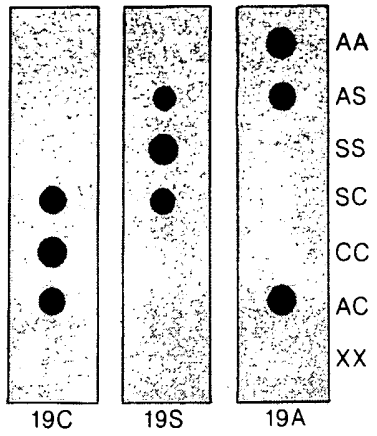


**Fig. 2** Genotype analysis of PCR-amplified genomic DNA using ASO probes. DNA was extracted from the buffy coat fractions of clinical blood samples of known β-globin genotype[11] and 1 μg portions of AA ($\beta^A\beta^A$), AS($\beta^A\beta^S$), SS ($\beta^S\beta^S$), SC ($\beta^S\beta^S$), CC ($\beta^C\beta^C$), AC ($\beta^A\beta^C$), and GM2064 (deletion mutant, see text) DNA were subjected to 25 cycles of PCR amplification as described previously[7]. Briefly, the samples were dispensed in 100 μl PCR reaction buffer (50 mM NaCl, 10 nM Tris, *p*H 7.6, 10 mM $MgCl_2$, 1 μM PC03, 1 μM PC04, 1.5 mM dATP, 1.5 mM dCTP, 1.5 mM dTTP, and 1.5 mM dGTP). After heating (10 min at 95 °C) to denature the DNA, the samples were transferred to a 30 °C heat block for 2 min to allow the primers to anneal to their target sequences. Primer extension was initiated by addition of one unit of Klenow fragment of *Escherichia coli* polymerase I (New England Biolabs) and incubated for 2 min at 30 °C. This cycle of heating, annealing, and extension was repeated 24 times; subsequent denaturations were by heating for 2 min. The final volume after 25 cycles was 150 μl. One-thirtieth of each reaction, (5 μl, containing 33 ng of the original genomic DNA), was adjusted to 0.4 N NaOH, 25 mM EDTA[17] in a 200 μl volume and applied to a Genatran-45 nylon filter (Plasco) with a Bio-Dot spotting apparatus (Bio-Rad). Three replicate filters were prepared. The ASO probes were phosphorylated at their 5′-termini with [γ-$^{32}$P]ATP and polynucleotide kinase and purified by spin dialysis[11]. The specific activities of the probes were between 3.5 and 4.5 μCi pmol$^{-1}$. Each filter was prehybridized individually in 8 ml 5×SSPE, 5× Denhardt, 0.5% SDS for 30 min at 55 °C. Probe (1 pmol) was then added and hybridized at 55 °C for 1 h. Filters rinsed twice in 2× SSPE, 0.1% SDS at room temperature, followed by a high stringency wash in 5×SSPE, 0.1% SDS for 10 min at 55 °C (for 19C) or 60 °C (for 19A and 19S) and autoradiographed for 2.5 h at −80 °C with a single intensification screen (Dupont Lightning Plus).

sequence[7]. The sequences of the 19-base ASO probes used here are identical to those described previously by Studencki *et al.*[8].

DNA was extracted from six blood samples from individuals with every possible diploid combination of the $\beta^A$, $\beta^S$, and $\beta^C$ alleles and from the cell line GMZ 2064, which has a homozygous deletion of the β-globin gene[9]. The DNA (1 μg from each sample) was subjected to 25 cycles of PCR amplification and 1/30th of the reaction product (33 ng) applied to a nylon filter as a 'dot blot'. Three replicate filters were prepared and each was hybridized with one of the three $^{32}$P-labelled ASOs under stringent conditions. The autoradiogram (Fig. 2) clearly indicates that each ASO annealed only to those DNA samples which contained at least one copy of the β-globin allele to which the probe was perfectly matched. For example, the $\beta^A$-specific probe, 19A, hybridized only to samples AA ($\beta^A\beta^A$), AS ($\beta^A\beta^S$), and AC ($\beta^A\beta^C$). The high degree of specificity contributed by PCR amplification is further emphasized by the lack of any detectable hybridization of the ASO probes to the GM2064 deletion mutant. The frequency of the specific β-globin target in PCR amplified DNA has been estimated by the analysis of cloned amplification products to be approximately 1%, an enrichment of $>10^5$ over unamplified genomic DNA[10]. It is this substantial reduction of complexity that allows the application of 19-base probes in a dot blot format and also makes possible the use of even shorter ASO probes capable of allelic discrimination under less stringent conditions (data not shown).

To estimate the minimum amount of DNA detectable, eight serial dilutions containing 1 to 128 ng of genomic AA DNA were amplified by 25 cycles and half of each sample was applied to a dot blot, then hybridized with the 19A probe. Equivalent dilutions of the 25-cycle amplifications of AA and SS DNA prepared in the previous experiment were included as controls. The autoradiogram (Fig. 3) reveals signals in all AA samples; the overall efficiency of amplification with nanogram quantities of template DNA is comparable to that obtained with 1 μg. Moreover, the subnanogram detection limit obtained with $^{32}$P-labelled oligonucleotide probes suggests that the analysis of nanogram quantities of genomic DNA with nonradioactive ASO probes (for example, biotin-labelled) should be feasible. Since one nanogram is the amount of genomic DNA present in only 150 diploid cells (assuming a haploid genome size of $3\times10^9$ bp) we considered the possibility that a crude lysate from a small number of cells would contain sufficiently little cellular debris to allow the DNA to serve as template for the PCR amplification procedure. Eight dilutions containing 75 to 9,600 cells from each of the cell lines MOLT-4 ($\beta^A\beta^A$)[11], SC-1 ($\beta^S\beta^S$)[11], and GM2064 were heated at 95 °C to lyse the cells and subjected to 25 cycles of amplification without purification. Equivalent amounts of purified MOLT-4 DNA were included as controls. The amplified samples were divided onto two dot blots and hybridized with the 19A and 19S probes. The autoradiogram (Fig. 4) shows a positive signal in an unfractionated cell lysate of only 75 cells. The efficiency of amplification appears to be only two- to three-fold less than with equivalent amounts of purified DNA. (MOLT-4 is reported to be 90% tetraploid (A. Green, personal communication) so that 150 cells would be expected to contain 2 ng of genomic DNA.) Both ASO probes retain allele specificity: 19A anneals only to MOLT-4 DNA and 19S only to SC-1 DNA. The variation from the expected two-fold reduction of signal intensities may reflect dilution artefacts associated with manipulating such small numbers of cells, and the significant reduction in signal intensity seen in both 9,600-cell samples may be due to inhibition of the amplification process by cellular

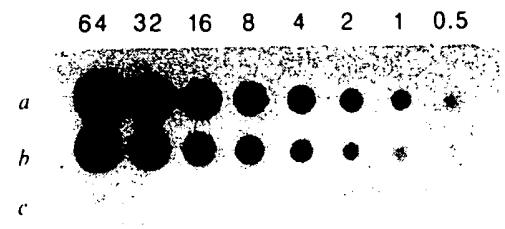


Fig. 3 PCR amplification of sub-microgram amounts of genomic DNA. *a*, Eight serial dilutions containing 128, 64, 32, 16, 8, 4, 2, and 1 ng of sample AA DNA were amplified by 25 cycles. One half of the amplification product (64 to 0.5 ng) was applied to a filter (row A). Controls consisted of equivalent dilutions of *b*, previously amplified AA and *c*, previously amplified SS DNA (Fig. 2). The filter was hybridized with the 19A probe and exposed for 6 h. See Fig. 2 (legend) for methods.

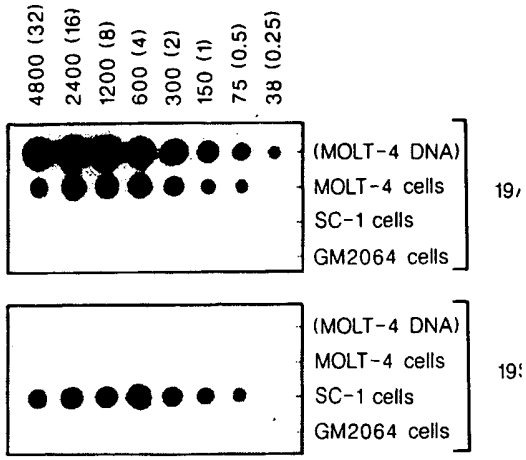


Fig. 4 Genotype analysis of PCR-amplified crude cell lysates using ASO probes. Cell concentrations of the cell lines MOLT-4 ($\beta^A\beta^A$), SC-1 ($\beta^S\beta^S$), and GM2064 (see text) were determined with a haemocytometer. Eight serial dilutions in PBS were prepared at $9.6\times10^5$, $4.8\times10^5$, $2.4\times10^5$, $1.2\times10^5$, $6.0\times10^4$, $3.0\times10^4$, $1.5\times10^4$, and $7.5\times10^3$ cells per ml. Ten μl of each dilution (containing 9,600, 4,800, 2,400, 1,200, 600, 300, 150 and 75 cells), were mixed with 35 μl $H_2O$ and heated for 10 min at 95 °C to lyse the cells. The lysates were added to 55 μl, 1.8 × PCR buffer (containing salts, primers, and dNTPs) and immediately subjected to 25 cycles of amplification. Equivalent amounts (64 to 0.5 ng) of purified MOLT-4 genomic DNA were also amplified. The PCR reaction products were divided onto two filters, probed with 19A and 19S and exposed for 6 h. Methods are as for Fig. 2 except that the amplification buffer included 10% DMSO (v/v), the annealing and extension reactions were performed at 37 °C instead of 30 °C[10] and the PCR amplification procedure was carried out in an automated device designed by the Cetus Instrumentation group.

debris. Use of unfractionated lysate reduces the time required for analysis to <8 hours.

PCR amplification and ASO hybridization have also been used to analyse genetic polymorphism at the HLA-DQα locus, which has localized allelic variability in the second exon (Fig. 5*a*). Using PCR primers complementary to conserved regions of this exon, a 242-bp fragment of DQα was amplified from DNA samples isolated from nine homozygous typing cell lines (HTC) and from two heterozygous cell lines, then hybridized with four different ASO probes (Figure 5*b*). The sequence of these probes was based on published data[12-15] as well as on our determinations of the sequences of complementary DNA, genomic and PCR-amplified clones (ref. 12 and Fig. 5, legend). Polymorphism in the HLA-DQ region, which consists of one α and one β chain locus, is detected serologically as the DQw1, DQw2, and DQw3 antigen series[16]. The DQA1 probe hybridizes selectively to amplified DNA from all cells typed as DQw1, which is associated with the DR1, DR2, and DRw6 serological types. As expected from the gene sequence, this probe also hybridizes to the DNA from the DR8 homozygous cell line TAB which is also DQw1. The DQA3 probe hybridizes to DNA from the DR3, DQw2 line AVL as well as to the unknown sequence in the DR5, DQw3 line JGL suggesting that these DQα alleles are homologous. Since DQA3 did not hybridize to the DNA from the other DQw3 lines (DKB and KOZ), there must be DQα sequence heterogeneity within this specificity. Thus the allelic variation in the HLA-DQα locus detected by this method appears to be greater than the polymorphic series defined by serological typing, and can also be more precisely localized. This approach is at present being applied to the analysis of polymorphism in the HLA-DQβ and DRβ loci.

With the use of nonisotopic probes and PCR automation (see Figs 4 and 5), this procedure, combining *in vitro* target

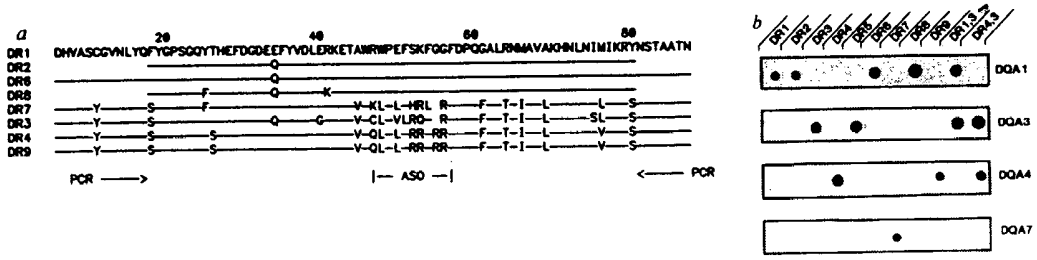


Fig. 5 Detection of allelic variation in HLA-DQα sequences. *a*, Allelic variation in the HLA-DQα gene; *b*, detection of HLA-DQα alleles. Protein sequences of the variable outer domain of the HLA-DQα chain, derived from cloned cDNA and PCR-amplified[10] sequences, are shown as differences from the sequence found in the DR1 (ref. 13) haplotype (upper line), and grouped to emphasize their similarities. The sequences from the DR1, DR2 (ref. 13) and DR6 (ref. 14) haplotypes (DQw1) are closely related, and those from the DR4 (ref. 14) and DR9 (ref. 16) haplotypes (DQw3) are identical. The region between positions 47 and 56 shows the greatest variation, and was therefore chosen as a target for four different ASO probes. In this segment, the DQα sequences from the DR1, 2, 6, and 8 (ref. 13) haplotypes are identical and homologous to the DQA1 ASO(5'-AACCTCCAAATTTGCTGAACTCAGGCCACCG-3'). The sequence from the DR7 (ref. 15) haplotype is homologous to the DQA7 ASO (TTCCACAGACTTAGATTTG) and the DR3 sequence (ref. 14) to the DQA3 probe (TGTTTGCCTGTTCTCAGAC). The sequences from the DR4 and DR9 haplotypes are identical and homogous to the DQA4 probe (TTCCGCAGATTTAGAAGAT). *b*, DNA was isolated[11] from nine homozygous typing cell lines and two heterozygous cell lines with HLA-DQα types DR1 (LG2), DR2 (PGF), DR3 (AVL), DR4 (DKB), DR5 (JGL), DR6 (APD), DR7 (LBF), DR8 (TAB), DR9 (KOZ), DR1,3 (B302), DR3,4 (B303). Each DNA sample (1 μg) was PCR-amplified for 28 cycles with the DQα primers as described (Fig. 4). The degree of amplification was determined by hybridization to a DQα cDNA probe labelled with $^{32}P$ by nick-translation, and equal amounts of product (2–10% of the final reaction volume) was denatured with two volumes of 0.5 M NaOH, 1.5 M NaCl. After neutralization with 0.5 M Tris, *p*H 7, 1.5 M NaCl to a final volume of 200 μl, the DNA was spotted in four replicates on MSI Nylon 66 (Fischer Scientific) and fixed to the filters by UV irradiation for 4 min. The filters were prehybridized for 1 h at 50° in 6 × SSPE, 10 × Denhardt, 0.5% SDS, and hybridized overnight with the ASO probes (at $2\times10^5$ cpm $ml^{-1}$) labelled with $^{32}P$ as described (Fig. 2), washed for 10 min at 37° in 0.1 × SSPE, 0.1% SDS, and autoradiographed for 90 min with an intensification screen.

amplification and ASO probes in a dot-blot format, promises to be a general and simple method for the detection of allelic variation. In addition, the ability to analyse genetic variation rapidly in minute amounts of purified DNA or cell lysates has important implications for a wide variety of genetic analyses.

We thank Corey Levenson, Lauri Goda and Dragan Spasic for the synthesis of oligonucleotides, Norm Arnheim for discussions, and Larry Johnson and Richard Leath for instrument and engineering support. We also thank John Sninsky for critical review, and Kathy Levenson, Eric Ladner, Tim Culp and Sharon Nilson for preparation of this manuscript.

*Note added in proof*: Recently, 50 ng samples of PCR-amplified genomic DNA have been analysed using a biotin-labelled ASO probe.

Received 30 May; accepted 5 September 1986.

1. Conner, B. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 278–272 (1983).
2. Kidd, V. J., Wallace, R. B., Itakura, K. & Woo, S. L. C. *Nature* **304**, 230–234 (1983).
3. Pirastu, M. *et al. New Engl. J. Med.* **309**, 284–287 (1983).
4. Wallace, R. B. *et al. Nucleic Acids Res.* **6**, 3543–3557 (1979).
5. Faloona, F. A. & Mullis, K. B. *Meth. Enzym.* (in the press).
6. Kan, Y. W. & Dozy, A. M. *Science* **209**, 388–391 (1980).
7. Saiki, R. K. *et al. Science* **230**, 1350–1354 (1985).
8. Studencki, A. B. *et al. Am. J. hum. Genet.* **37**, 42–51 (1985).
9. Tuan, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 6937–6941 (1983).
10. Scharf, S., Horn, G. & Erlich, H. A. *Science* **223**, 1076–1078 (1986).
11. Saiki, R. K., Arnheim, N. & Erlich, H. A. *Biotechnology* **3**, 1008–1012 (1985).
12. Bugawan, T., Horn, G., Long, C. & Erlich, H. A. (Manuscript in preparation).
13. Horn, G. *et al.* (manuscript in preparation).
14. Chang, H., Moriuchi, T. & Silver, J. *Nature* **305**, 813–815 (1983).
15. Moriuchi, J., Moriuchi, T. & Silver, J. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3420–3424 (1985).
16. Albert, E. D., Baur, M. P. & Mayr, W. R. *Histocompatibility Testing*, 4–8 (Springer, Berlin, 1984).
17. Reed, K. C. & Mann, D. A. *Nucleic Acids Res.* **13**, 7207–7221 (1985).

# Common evolutionary origin of legume and non-legume plant haemoglobins

**Jörg Landsmann, Elizabeth S. Dennis, Thomas J. V. Higgins, Cyril A. Appleby, Alexander A. Kortt* & W. James Peacock**

Commonwealth Scientific and Industrial Research Organisation, Division of Plant Industry, GPO Box 1600, Canberra, ACT 2601, Australia
* Division of Protein Chemistry, CSIRO, Parkville, Victoria 3052, Australia

**The detection of haemoglobins in distantly-related non-legume plant families[1,2] as well as in the legume families raises the question of whether the origin of the plant haemoglobins was single or multiple. We have isolated a haemoglobin gene from *Parasponia andersonii* (Ulmaceae) which shows more than 50% nucleotide sequence homology with the haemoglobin genes of legume plants and has three introns at identical positions to the leghaemoglobin introns[3–5]. The *Parasponia* gene has homology to haemoglobin genes in another distantly-related nodulating plant, *Casuarina*. It also hybridizes at high stringency to sequences in a related, but non-nodulating genus, *Trema*. We conclude that the globin gene family may be widespread in modern plants, that plant haemoglobins may have a cryptic function in non-symbiotic tissue and that plant haemoglobins have evolved by vertical descent, probably from an ancestor common to modern plants and animals.**

A number of plants outside the legume families are known to form nitrogen-fixing root nodules in symbiotic association with *Rhizobium*[6] or actinomycetes[7]. Haemoglobin has been detected in nodules of several non-legume plants[1,2] and characterized from *Parasponia* (Ulmaceae)[8,9] and *Casuarina* (Casuarinaceae)[10]. Modern systems of plant classification show the known nodulating, haemoglobin-containing species to be widely distributed over several superorders of dicotyledons[7,11,12] (Fig. 1). No continuous evolutionary pattern is evident.

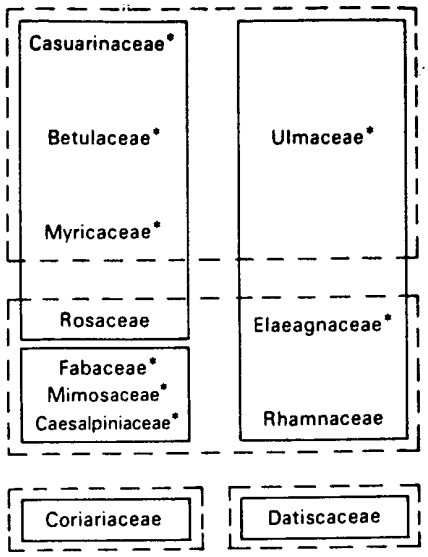


**Fig. 1** Phylogenetic relationship of dicotyledonous plant families in which some species have been shown to form nitrogen-fixing nodules[7]. *Parasponia* is the only genus of the Ulmaceae which has been shown to nodulate, whereas in the legume families nearly all genera of the Fabaceae and Mimosaceae nodulate; the Caesalpiniaceae are mostly non-nodulating. Families in which at least one species has been shown to contain nodule haemoglobin[1,2] are marked with an asterisk. Solid line, superorders according to Dahlgren[12]; dotted line, according to Cronquist[11]. Note that these superorders also contain ~150–250 (depending on the classification system) non-nodulating families not listed here. The $N_2$-fixing symbiosis in legumes is always with *Rhizobium* bacteria, *Parasponia* is the only known non-legume genus to be nodulated by *Rhizobium* strains, some of which also infect legumes[6]. Other non-legumes undergo symbiosis with *Frankia*, an actinomycete[7].

Although there is 40% protein homology between soybean and *Parasponia* haemoglobin[9], a soybean leghaemoglobin complementary DNA (cDNA) clone did not hybridize to *Parasponia* DNA or nodule RNA, even at reduced stringency. A 17mer and a 14mer oligonucleotide sequence deduced from the *P. andersonii* haemoglobin I protein sequence[9] were made (see Fig. 4). Both oligonucleotides hybridized to an RNA 750 bases long in Northern blots of *Parasponia* nodules (Fig. 2*a*). The 17mer was used as primer for synthesis of radioactive cDNA, which was used as a probe; it also hybridized to a 750-base RNA in Northern blots (Fig. 2*b*) and identified haemoglobin-specific clones in a nodule cDNA library. Two positive clones were analysed; pL209 contained the entire coding region for *Parasponia* haemoglobin I whereas pL201 lacked the first 16 codons. Using pL209 as a probe, three independent genomic clones were isolated from a *Bam*HI–EMBL4 genomic library. All three clones contained one complete *Parasponia* haemoglobin gene on a 4.8 kilobase (kb) *Eco*RI end fragment within a 13-kb *Bam*HI fragment. The sequences of the genomic clones, pL305 and pL401, correspond to the cDNA clones pL209 and pL201 respectively (Fig. 3).

The coding region of the *Parasponia* haemoglobin gene spans 486 nucleotides which translate into 161 amino acids excluding the N-terminal Met (Fig. 4). Predictions from the DNA sequence showed discrepancies with the published protein sequence[9]. Resequencing of the relevant peptides confirmed an additional Phe at position 36, an additional Val at position 73 and a Cys at position 78 instead of the previously reported Thr. Four additional amino acids (Ser-Ser-Ser-Glu) were found at the N-terminus (Fig. 4). *Parasponia* haemoglobin has the widely conserved globin residues C2-Pro, CD1-Phe, proximal His (E7) and distal His (F8) and, as in other plant haemoglobins, an extended EF-region[13]. The protein coding DNA region of the *Parasponia* haemoglobin gene shows more than 50% homology